

A global sensitivity analysis approach for marine biogeochemical modeling

Clémentine Prieur, Laurence Viry, Eric Blayo, Jean-Michel Brankart

► To cite this version:

Clémentine Prieur, Laurence Viry, Eric Blayo, Jean-Michel Brankart. A global sensitivity analysis approach for marine biogeochemical modeling. Ocean Modelling, Elsevier, 2019, 139 (101402), pp.1-38. 10.1016/j.ocemod.2019.101402 . hal-01952797v2

HAL Id: hal-01952797

<https://hal.inria.fr/hal-01952797v2>

Submitted on 17 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A global sensitivity analysis approach for marine biogeochemical modeling

C. Prieur^a, L. Viry^a, E. Blayo^a, J.-M Brankart^b

^a*Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP*, LJK, 38000 Grenoble, France*

^b*Univ. Grenoble Alpes, CNRS, Grenoble INP*, IGE, 38000 Grenoble, France*

** Institute of Engineering Univ. Grenoble Alpes*

Abstract

This paper introduces the Sobol' indices approach for global sensitivity analysis (SA), in the context of marine biogeochemistry. Such an approach is particularly well suited for ocean biogeochemical models, which make use of numerous parameters within large sets of differential equations with complex dependencies. This SA allows for a detailed study of the relative influence of a large number of input parameters on output quantities of interest to be chosen. It is able to distinguish between direct effects of these parameters and effects due to interaction between two or more parameters. Although demanding in terms of computation, such a tool is now becoming affordable, thanks to the development of distributed computing environments. An applicative example is presented with the MODECOGeL biogeochemical model, and illustrates the advantages of this approach over standard local SA.

Keywords: sensitivity analysis, Sobol' indices, marine biogeochemistry

1. Introduction

Marine biogeochemical models are now commonly included as modules in complex ocean circulation modeling systems. They are thus increasingly used for many applications. However, the use of these models raises difficult questions, especially regarding their tuning. They are generally systems of nonlinear ordinary differential equations, with each equation expressing the time evolution of a given state variable due to hydrodynamical effects (transport and diffusion) and to fluxes between the various components of the ecosystem. Two features of these models are noteworthy. First, the evaluation of the fluxes involves a wide variety of processes, and hence numerous

parameters (typically several times more than the number of variables). Secondly, the values of these parameters are often quite poorly known. In reality, the values of the parameters depend on the physical and biogeochemical context, while in practice the available reference values were usually estimated only in a particular situation (e.g. during a field experiment or in laboratory experiments), not in the actual context of interest. The uncertainty of these values is thus generally quite large (see for example Schartau et al. (2017) for a review on the identification of such parameters).

In this context, *sensitivity analysis* (SA), i.e. methods that aim to quantify the relative influence of the inputs on some given outputs in a complex system like a numerical model, are a valuable tool. Indeed, they can help better understand the model itself, and identify which parameters are most influential and should be calibrated carefully. These methods may be divided in two main categories: *local* sensitivity methods that consider the behavior of the solution with respect to small parameter variations and *global* sensitivity methods that determine the behavior of the solution under parameter perturbations of arbitrary magnitude. A common approach is to conduct a few experiments in which the values of parameters vary, either one-at-a-time (OAT), or simultaneously (e.g. Druon and Le Fèvre (1999); Baklouti et al. (2006); Kriest et al. (2012)). Some additional techniques are also sometimes used, such as linear error propagation (Omlin et al., 2001) or the Gaussian emulation machine approach (Scott et al., 2011), which clearly involve global sensitivity analysis.

Another approach defines the sensitivity of the output with regard to the input as the corresponding gradient (both input and output variables must of course be continuous real-valued quantities). This so-called *gradient based SA* thus requires computing gradients. This is straightforward for simple cases, but challenging in general. For instance, the gradient of a single output quantity Q with regard to a constant parameter P can be approximated by $(Q(p + \alpha) - Q(p))/\alpha$, where p is the current value of P and α is taken to be small. The computation of this approximate gradient simply requires running the model twice, using successively p and $p + \alpha$ for parameter P . However estimating the gradient can be much more difficult if P is multivariate or non-constant (e.g. if P is a space and/or time dependent coefficient). In such cases, the preceding approach of computing growth rates requires $N + 1$ evaluations of the model, where N is the dimension of P (e.g. the number of space-time grid points). If N is large, an alternative to computing growth rates is to use the so-called *adjoint method*, which provides the exact gradient.

49 This approach was used for instance in Fennel et al. (2001), Faugeras et al.
50 (2003) and Tjiputra et al. (2007).

51 However the gradient is computed, it is important to understand that
52 gradient based SA is a local method, in the sense that the gradient is a local
53 notion, computed in the vicinity of the current value of P . Therefore, this
54 gradient, which is a way to quantify the influence of P on Q , can be quite
55 different depending on the value of P chosen. This limitation can be avoided
56 by using an approach based on a global sensitivity measure, i.e. an approach
57 that quantifies the influence of P on Q taking into account the possible vari-
58 ation of P . Such a quantification is provided, for example, by *Sobol' indices*,
59 which will be introduced in the next section. Given their global character,
60 computing such indices may of course require a huge number of model evalu-
61 ations, which increases with the input space dimension. Therefore it is usual
62 to first carry out, as a preliminary step, a screening analysis, such as the OAT
63 screening approach introduced in Morris (1991). A screening procedure aims
64 at fixing the input parameters whose influence on the output is qualified as
65 negligible after a rough exploration of the input parameter space. It reduces
66 the input space dimension. Then, in a second step, a Monte Carlo sampling-
67 based sensitivity and ranking analysis is run (see, e.g., Sankar et al., 2018) or
68 Sobol' indices are computed (see, e.g., Morris et al., 2014; Wang et al., 2018).
69 However, given the increases in computational resources, we think that per-
70 forming a direct global SA is now feasible, even for a high-dimensional input
71 space, and that it is time for the scientific community to make computa-
72 tion of Sobol' indices the first choice, particularly when dealing with highly
73 parameterized models.

74 In this context, the aim of this paper is to introduce the Sobol' indices
75 approach for global sensitivity analysis, and to illustrate its feasibility and its
76 scientific relevance in the context of marine biogeochemistry. Note, however,
77 that our focus is to present the methodological tools in the context of a par-
78 ticular example, rather than to conduct an in-depth physical analysis. First,
79 the global SA approach and the Sobol' indices are introduced in Section 2.
80 Then, in Section 3, we present the biogeochemical model, its many uncer-
81 tain parameters and the selected output quantities. Section 4 is devoted to
82 implementation aspects of the SA, and Section 5 focuses on results that help
83 illustrate the features of this method. In particular, a comparison with a
84 gradient-based local analysis is presented.

85 2. Global sensitivity analysis: the framework

As indicated above, the aim of sensitivity analysis is to determine which model inputs are the most influential on some given model outputs. In the following, a model output y is considered to be a deterministic scalar function of some model inputs $\mathbf{x} = (x_1, \dots, x_d)$, where these inputs belong to a domain Δ :

$$y = y(\mathbf{x}) = y(x_1, \dots, x_d) \quad \text{with } \mathbf{x} \in \Delta$$

86 In Section 3, we consider various scalar outputs, corresponding to different
 87 quantities of interest. In this paper, we adopt the stochastic framework of
 88 global sensitivity analysis. Unlike local sensitivity analysis, which analyses
 89 how a small perturbation near an input space value $x^0 = (x_1^0, \dots, x_d^0)$ in-
 90 fluences the value of the scalar output, global sensitivity analysis considers
 91 the whole variation range in the input space. More precisely, each input
 92 parameter is considered to be a random variable X_j ($j = 1, \dots, d$), where
 93 its uncertainty is modeled by some one-dimensional probability distribution.
 94 The output Y is then assumed to be a scalar random variable $Y = f(\mathbf{X})$,
 95 where $\mathbf{X} = (X_1, \dots, X_d)$ with the X_i assumed to be independent. Various
 96 sensitivity measures have been proposed in the global framework. We focus
 97 in this paper on variance-based sensitivity measures, introduced in Sobol'
 98 (1993).

99 2.1. Variance-based sensitivity measures

100 In order to quantify the influence of the variations of X_j on the variations
 101 of Y , let us consider the conditional expectation $E(Y|X_j = x_j)$. It corre-
 102 sponds to the mean value of Y over the probability distributions of the X_k
 103 ($k \neq j$), when X_j is fixed to x_j . The corresponding random variable, when
 104 considering the variations of X_j , is $E(Y|X_j)$, and its variance quantifies the
 105 influence of X_j on the dispersion of Y . The so called Sobol' sensitivity indices
 106 are obtained by normalizing this variance by the total variance of the output
 107 Y , which is assumed to be finite and non null. Thus the first-order Sobol'
 108 sensitivity index of input parameter X_j is defined as

$$S_{\{j\}} = \frac{\text{Var}(E(Y|X_j))}{\text{Var}(Y)}. \quad (1)$$

109 It belongs to the interval $[0, 1]$.

110 *2.2. The Sobol' decomposition*

111 More generally, starting from the functional Analysis of Variance (ANOVA)
 112 decomposition (Hoeffding, 1948; Efron and Stein, 1981; Owen, 1992; Sobol',
 113 1993), one can define sensitivity indices of any order $r \in \{1, \dots, d\}$. Let us
 114 first introduce some notation. We assume that f is a real square integrable
 115 function, \mathbf{u} is a subset of $\{1, \dots, d\}$, \mathbf{u}^c stands for its complement, its car-
 116 dinality is denoted by $r = |\mathbf{u}|$, and $\mathbf{X}_{\mathbf{u}}$ represents the random vector with
 117 components X_j , $j \in \mathbf{u}$. The functional ANOVA decomposition then states
 118 that $Y = f(\mathbf{X})$ can be uniquely decomposed into summands of increasing
 119 size

$$Y = f(\mathbf{X}) = \sum_{\mathbf{u} \subseteq \{1, \dots, d\}} f_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}) \quad (2)$$

120 where $f_{\emptyset} = \mathbb{E}[Y]$ and the other components have zero mean value and are
 121 mutually uncorrelated. This decomposition (2) yields a corresponding de-
 122 composition of the variance of the output:

$$\text{Var}(Y) = \sum_{\mathbf{u} \subseteq \{1, \dots, d\}, \mathbf{u} \neq \emptyset} \text{Var}(f_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})) . \quad (3)$$

123 For any $j \in \{1, \dots, d\}$, the term $\text{Var}(f_{\{j\}}(X_j))$ corresponds to the part of
 124 the output variance explained by parameter X_j . For any $j, j' \in \{1, \dots, d\}$,
 125 $j < j'$, the term $\text{Var}(f_{\{j, j'\}}(X_j, X_{j'}))$ corresponds to the part of the output
 126 variance explained by combined effects of parameters X_j and $X_{j'}$. More
 127 generally, for any $\mathbf{u} \subseteq \{1, \dots, d\}$, the term $\text{Var}(f_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}))$ corresponds to the
 128 part of the output variance explained by combined effects of parameters X_j ,
 129 $j \in \mathbf{u}$. Then, for any $\mathbf{u} \subseteq \{1, \dots, d\}$, the Sobol' index (Sobol', 1993) of order
 130 $r = |\mathbf{u}|$ associated to the vector $\mathbf{X}_{\mathbf{u}}$ is defined as

$$S_{\mathbf{u}} = \frac{\sigma_{\mathbf{u}}^2}{\sigma^2} = \frac{\text{Var}(f_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}))}{\text{Var}(Y)} . \quad (4)$$

131 The main effect of the j^{th} factor is thus measured by $S_{\{j\}}$. Then the effect
 132 due to the specific interaction between the j^{th} and k^{th} factors ($k \neq j$) is
 133 measured by $S_{\{j, k\}}$. And so on for higher order indices (Saltelli et al., 2000).
 134 Corresponding cumulative effects are defined by so-called closed Sobol' in-
 135 dices, defined as

$$S_{\mathbf{u}}^{\text{closed}} = \sum_{\mathbf{v} \subseteq \mathbf{u}} S_{\mathbf{v}} . \quad (5)$$

136 We thus have for instance $S_{\{j,k\}}^{\text{closed}} = S_{\{j\}} + S_{\{k\}} + S_{\{j,k\}}$, which means that
 137 the closed second-order index $S_{\{j,k\}}^{\text{closed}}$ measures the effect of the interactions
 138 between parameters X_j and X_k in addition to the main effect of each of these
 139 two parameters.

140 For any $j \in \{1, \dots, d\}$, we also define a total sensitivity index $S_{\{j\}}^{\text{tot}}$
 141 (Homma and Saltelli, 1996) to express the overall sensitivity to an input
 142 X_j as

$$S_{\{j\}}^{\text{tot}} = \sum_{\mathbf{v} \subset \{1, \dots, d\} \text{ with } j \in \mathbf{v}} S_{\mathbf{v}}. \quad (6)$$

143 We have then, for instance, in the case where $d = 3$: $S_{\{1\}}^{\text{tot}} = S_{\{1\}} + S_{\{1,2\}} +$
 144 $S_{\{1,3\}} + S_{\{1,2,3\}}$.

145 2.3. Estimating the Sobol' indices

146 For simple models, sensitivity analysis can sometimes be done analyti-
 147 cally, by direct examination of their mathematical expression. However, this
 148 is of course generally not the case for complex models. In those cases, one
 149 evaluates the model for selected values of the input parameters, and the re-
 150 sulting output values are used to estimate sensitivity indices of interest. For
 151 interested readers, Monte Carlo based procedures for the estimation of Sobol'
 152 indices are described in detail in Appendix A.

153 Let us summarize the different strategies we apply in Section 5:

- 154 • In Subsection 5.1, we apply the replication procedure introduced in
 155 Mara and Rakoto Joseph (2008) (and further studied in Tissot and
 156 Prieur (2015)) to estimate all first-order Sobol' indices with only two
 157 replicated d -dimensional Latin hypercube samples of size n , that is with
 158 only $2n$ model evaluations.
- 159 • In Subsection 5.2, we apply the replication procedure introduced in Tis-
 160 sot and Prieur (2015) to estimate all closed second-order Sobol' indices
 161 with only two replicated d -dimensional randomized orthogonal arrays
 162 of strength 2 and size n , that is again with only $2n$ model evaluations.
 163 Due to constraints in the construction of orthogonal arrays of strength
 164 2, n must be chosen as q^2 , with q a prime number greater than, or equal
 165 to, $d - 1$.
- 166 • In Subsection 5.3, we apply the procedure introduced in (Saltelli, 2002,
 167 Theorem 1) to estimate all first-order and total Sobol' indices with a

168 cost of only $(d + 2)n$ model evaluations. This procedure is based on
169 combinatorial arguments.

170 These procedures are implemented in the functions `sobolroalhs`, `sobolSalt`
171 of the package `sensitivity`. All these strategies are detailed in Appendix
172 A. A crucial point in the above summary is that the cost (in terms of number
173 of model evaluations) required to estimate total Sobol' indices is only linear
174 in the input space dimension d . In the following (see Section 4), we present
175 an implementation using a grid computing environment. However, even in
176 such a framework, the estimation of total Sobol' indices be expensive if d
177 is large. Thus, depending on the computational resources of the study, one
178 could avoid the estimation of these total indices and focus instead on the
179 estimation of all first-order and closed second-order Sobol' indices as a first
180 step in the sensitivity analysis.

181 3. Description of the ocean biogeochemical model

182 The SA approach is applied to the 1D biogeochemical MODECOGeL
183 model of the ocean mixed layer described in subsection 3.1. The uncer-
184 tain input parameters of the model (vector \mathbf{x} in Section 2) are described in
185 subsection 3.2 and several output quantities (y in Section 2) are listed in
186 subsection 3.3, corresponding to key model results whose sensitivity to the
187 input parameters is not obvious.

188 3.1. The MODECOGeL model

189 The model used in this paper is MODECOGeL¹. It was developed for in-
190 vestigating the biogeochemical activity in the Ligurian sea by Lacroix (1998)
191 by coupling a 1D hydrodynamic model of the mixed layer to a 12-component
192 ecosystem model.

193 The hydrodynamic model is a 1D version of the GHER primitive equa-
194 tions model (Nihoul and Djenidi, 1987). The state variables are the horizon-
195 tal velocity, the potential temperature, the salinity, and the turbulent kinetic
196 energy. A full description of the model can be found in Lacroix and Nival
197 (1998) or Lacroix and Grégoire (2002), where it is applied to simulate the
198 behavior of the system during the FRONTAL oceanographic campaigns from

¹MODECOGeL: MODèle d'ECOsystème du GHER (GeoHydrodynamics and Environ-
ment Research) et du LOV (Laboratoire d'Océanographie de Villefranche-sur-Mer)

199 1984 to 1988. In the present paper, the model is applied to years 2006–2007,
 200 and the atmospheric dataset is extracted mainly from the Côte d’Azur me-
 201 teorological buoy located at the DYFAMED station (BOUSSOLE project)
 202 at hourly frequency (Marty and Chiavérini, 2010).

203 The ecosystem model provides a 12-component description of the ecosys-
 204 tem of the Ligurian Sea (see state variables in Table 1). A detailed descrip-
 205 tion of this model can be found in Lacroix and Grégoire (2002). The time
 206 evolution of each state variable is governed by the equation:

$$\frac{\partial C_i}{\partial t} = \text{ADV}_i + \text{DIFF}_i + \text{SMS}_i \quad \text{with} \quad \text{SMS}_i = \sum_{j \neq i} \text{FLUX}(C_j \rightarrow C_i) \quad (7)$$

207 where ADV_i and DIFF_i are advection and diffusion terms (governed by the hy-
 208 drodynamic model), and SMS_i is the “source minus sink” term summing up
 209 the fluxes ($\text{FLUX}(C_j \rightarrow C_i)$) between the various components of the ecosystem
 210 (conservation of course imposes that $\text{FLUX}(C_j \rightarrow C_i) = -\text{FLUX}(C_i \rightarrow C_j)$).
 211 These fluxes can be sorted into several categories: primary production, sec-
 212 ondary production, mortality, exudation, excretion, growth of bacteria, de-
 213 composition of particulate organic matter, and nitrification. We refer to Ap-
 214 pendix B for a detailed description of these different processes.

215 Equation (7) is solved numerically between the sea surface and 405 m
 216 depth using a constant vertical discretization (1 m) and a constant time step
 217 (6 minutes). Outputs are saved daily at all depths.

218 It is important to note that the mathematical expression of these flux
 219 terms contains numerous parameters whose values are not known precisely.
 220 In the following, consistent with the objective of this study, we provide only
 221 a brief overview of these model parameters, with a specific focus on the
 222 assumptions that we make regarding their respective uncertainties.

223 3.2. Model parameters

224 The biogeochemical fluxes ($C_j \rightarrow C_i$) parameterized in MODECOGeL
 225 are summarized in Table B.8 (Appendix B). Each flux depends on several of
 226 the parameters listed in Table C.9 (Appendix C). Moreover, the parameteri-
 227 zation of each process as a joint function of the model state and parameters is
 228 often complex and nonlinear. As a result, it is usually impossible for the user
 229 to know the sensitivity of the whole system to the parameters. A systematic
 230 computational approach is thus needed.

Variable	Acronym	Name
C_1	NO3	Nitrate
C_2	NH4	Ammonium
C_3	PicP	Picophytoplankton
C_4	NanP	Nanophytoplankton
C_5	MicP	Microphytoplankton
C_6	NanZ	Nanozooplankton
C_7	MicZ	Microzooplankton
C_8	MesZ	Mesozooplankton
C_9	BAC	Bacteria
C_{10}	DON	Dissolved organic nitrogen
C_{11}	POM1	Particulate organic matter (size 1)
C_{12}	POM2	Particulate organic matter (size 2)

Table 1: Model state variables.

231 To apply the SA method described in Section 2, a probability distribution
232 must be specified for each input parameter. This has been done here using
233 the following guidelines:

- 234 • In the absence of any reliable information about possible correlations,
235 the uncertainties of the various parameters are assumed independent.
- 236 • Most parameters are constrained to be either positive or negative, for
237 consistency with the formulations used to parameterize the processes.
238 They are assumed to follow a Gamma distribution.
- 239 • Parameters constrained between 0 and 1 are assumed to follow a Beta
240 distribution.
- 241 • Some parameters are constrained to be larger than 1. Their logarithm
242 is assumed to follow a Gamma distribution.
- 243 • Some parameters are not constrained, and are assumed to follow a
244 Gaussian distribution.
- 245 • Three different values for standard deviations are used (5%, 20%, 50%
246 of the expected value) according to the confidence we have in the pa-
247 rameters. These values were provided by biogeochemical modelers us-
248 ing a priori knowledge.

249 The resulting probability distributions are given in Table C.9 (Appendix C).

250 3.3. Quantities of interest

251 The quantities of interest (QoI), i.e. the output values y , must be defined
 252 according to the main scientific objectives of the sensitivity study. In the
 253 present case, for this example, we have chosen to focus on characterizing the
 254 simulation of phytoplankton (concentrations C_3, C_4, C_5 in Table 1), which
 255 is at the base of the marine food web. As an additional quantity, we also
 256 introduce chlorophyll concentration (noted C_0), which is what is observed
 257 by ocean color data, and which can be approximately computed from phyto-
 258 plankton concentrations using a constant chlorophyll to nitrogen ratio (α):

$$C_0 = \alpha(C_3 + C_4 + C_5) \quad (8)$$

259 Since sensitivity analysis applies to scalar output quantities, to apply the SA
 260 method we need to reduce the time and space variations of C_0, C_3, C_4, C_5 to
 261 some scalar indicators. Note that the cost of the method is almost indepen-
 262 dent of the number of these indicators. However, they must be defined before
 263 running the sensitivity study. To illustrate the method, we thus decided to
 264 introduce a range of different QoI characterizing C_0, C_3, C_4, C_5 , without lim-
 265 iting our choice to simple linear diagnostics.

266 Table 2 summarizes the QoI Y_{ij} that we will use in our application. The
 267 second index j corresponds to the computed diagnostic while the first index
 268 i corresponds to the concentration ($i = 0, 3, 4, 5$) to which it is applied. This
 269 set of five diagnostics is meant to characterize (i) the maximum intensity of
 270 the phytoplankton spring bloom, (at the surface and as a vertical average),
 271 (ii) the time at which it occurs, and (iii) the overall average over the whole
 272 simulation.

273 4. Practical aspects of the sensitivity analysis

274 In this section we provide more details about the numerical implementa-
 275 tion of the sensitivity analysis, i.e. the computation of estimates of all first,
 276 closed second-order and total Sobol' indices.

277 4.1. Grid computing environment

278 In most cases, the primary cost of the global sensitivity analysis is due
 279 to the need for numerous evaluations of the model. In our study each model

Index j	Diagnostic	Definition
1	surface maximum	$\max_t C_i(0, t)$
2	time of surface maximum	$\operatorname{argmax}_t C_i(0, t)$
3	maximum of vertical average	$\max_t \frac{1}{Z} \int_0^Z C_i(z, t) dz$
4	time of maximum of vertical average	$\operatorname{argmax}_t \frac{1}{Z} \int_0^Z C_i(z, t) dz$
5	time and vertical average	$\frac{1}{ZT} \int_0^T \int_0^Z C_i(z, t) dz dt$

Table 2: Quantities of interest Y_{ij} . The maximum depth for averaging is $Z = 40$ m and T is the total duration of the experiment.

280 evaluation is short (approximately 50s). However, $2n$ evaluations are required
281 to estimate all first-order or all closed second-order Sobol’ indices, and $(d +$
282 $2)n$ evaluations are required to estimate all total Sobol’ indices, where n is
283 typically of order $10^3 - 10^6$. Furthermore, we must manage the model’s input
284 and output files. This problem thus requires specific tools to automatically
285 organize data management and access to computational resources. Note
286 also that, since we have to perform a very large number of model runs, our
287 computing environment must include fault tolerance.

288 A grid computing environment (i.e. a distributed architecture of com-
289 puters linked by communication networks and managed by “middleware”
290 software) is particularly well suited for this kind of application, in which
291 multiple parallel computations take place independently without the need to
292 communicate intermediate results between processors. In our implementa-
293 tion we used computing resources at the University of Grenoble², distributed
294 on three sites (10 clusters, 6600 CPU cores and 176 GPU cores). These re-
295 sources are connected by a local grid³ associated with several storage nodes
296 distributed among the three sites as close as possible to each supercomputer,
297 and managed by the middleware iRods⁴. Each cluster uses the local resource
298 manager OAR⁵ and accesses the full 6600 CPU cores through the middle-

²see <https://ciment.ujf-grenoble.fr/wiki-pub/index.php>

³https://ciment.ujf-grenoble.fr/wiki-pub/index.php/Grid_computing

⁴see irods.org

⁵see oar.imag.fr

299 ware CIGRI⁶ which launches embarrassingly parallel jobs on idle processors.
 300 CIGRI is able to manage job failures in a smart way, which allows the user
 301 to submit many small jobs. As a final step, we also need to transfer many
 302 input and result files distributed over different iRods resources on a laptop or
 303 a computer server of a laboratory. This is achieved efficiently using a python
 304 script.

305 4.2. Implementation of our study

306 As indicated above, the MODECOGeL model presented in Section 3 in-
 307 cludes 74 uncertain input parameters, each with an associated probability
 308 distribution (see Table C.9). The global sensitivity analysis is done with the
 309 R package **sensitivity** (Pujol et al., 2017) (R⁷ is a free software environ-
 310 ment for statistical computing and graphics).

311 The design of the experiments (DoE) and the estimation of the Sobol’ in-
 312 dices and their associated confidence intervals use the functions **sobolroalhs**
 313 and **sobolSalt**. Function **sobolroalhs**, which estimates all first-order and
 314 all closed second-order indices, is based on the replication procedure briefly
 315 described in Subsection 2.3 and detailed in Appendix A. It makes use of
 316 four replicated designs of size n (two for first-order indices and two for closed
 317 second-order indices).

318 The grid deployment used for the estimation of all first-order Sobol’ in-
 319 dices is done as follows (see also Figure 1):

- 320 • Construct two replicated Latin Hypercube Samplings (LHS) of size n
 321 with the **sobolroalhs** function.
- 322 • Evaluate the model on the DoE. In order to minimize the overhead
 323 corresponding to the submission of each evaluation and to optimize the
 324 use of the “best effort” mode of the batch manager, the model runs
 325 are performed one hundred at a time. Each group of simulations cor-
 326 responds to 100 different sets of the 74 input parameters, and requires
 327 an average cpu time of 84 minutes and a small input and output files
 328 of approximately 200Ko. For $n = 10^6$, we thus submitted 20 000 jobs
 329 on the computing grid.

⁶<http://ciment.ujf-grenoble.fr/cigri/dokuwiki/doku.php>

⁷<http://cran.r-project.org>

- Merge the $2n$ files containing the evaluations of the quantities of interest on the DoE.
- Treat outliers. A small percentage of the model evaluations fail or lead to completely spurious results, due to unrealistic combinations of the input parameters. The quantities of interest are then set to NaN and will be treated as missing values in the computation of Sobol' indices.
- Compute Sobol' indices using the function `sobolroalhs`.

Note that the experimental designs and the functions used to compute Sobol' indices differ for the estimations of closed second-order and total indices, as detailed in Section 2.3.

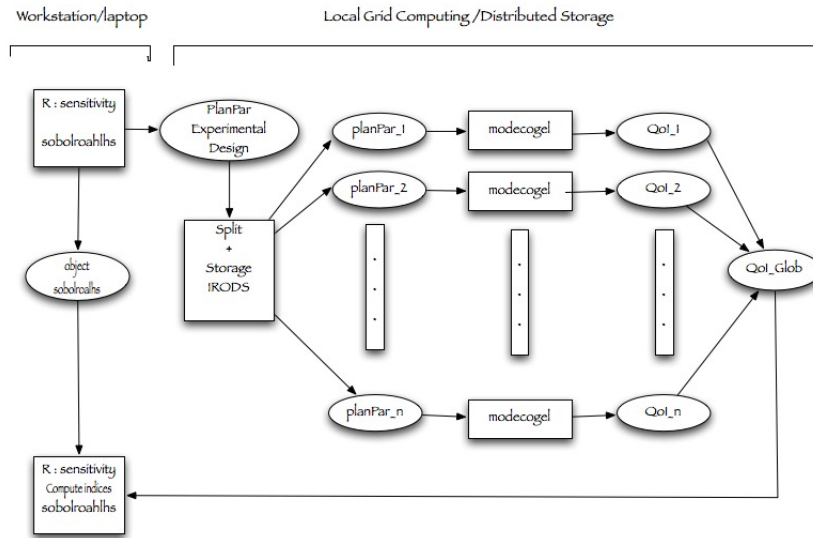


Figure 1: The steps for the estimation of all first-order (or all closed second-order) Sobol' indices with the `sobolroalhs` function of the R `sensitivity` package. The experimental design `PlanPar` is split into p sets of simulations (100 simulations each in our case). Each set of simulations is performed using `MODECOGeL` and the `QoI` are computed for each simulation. All values for the `QoI` are grouped in a single file `QoIGlob`, which is sent to `sobolroalhs` for the actual computation of the Sobol' indices.

5. Results

We focus here on the quantities of interest (QoI) which are summarized in Table 3 and described in Section 3. Once again, our aim with this global sensitivity analysis of the MODECOGeL model is not to perform an in-depth biogeochemical analysis of the results (this is outside the scope of this study), but rather to show how these statistical tools provide a better understanding of complex systems involving many parameters.

Quantity of interest	Y_{ij}	Description
maxc	Y_{01}	annual maximum of surface chlorophyll concentration
timechl	Y_{02}	time of maximum of surface chlorophyll concentration
moyc	Y_{05}	time and vertical average of chlorophyll concentration
maxpp	Y_{31}	maximum of surface picophytoplankton concentration
timepp	Y_{32}	time of maximum of surface picophytoplankton concentration
maxmoypp	Y_{33}	maximum of vertical average of picophytoplankton concentration
timemoypp	Y_{34}	time of maximum of vertical average of picophytoplankton concentration
maxnp	Y_{41}	surface maximum of nanophytoplankton concentration
timenp	Y_{42}	time of maximum of surface nanophytoplankton concentration
maxmoynp	Y_{43}	maximum of vertical average of nanophytoplankton concentration
timemoynp	Y_{44}	time of maximum of vertical average of nanophytoplankton concentration
moynp	Y_{45}	time and vertical average of nanophytoplankton concentration

Table 3: Quantities of interest analyzed in the present work. See also §3.3 and Table 2 for the explanation of the notation Y_{ij}

5.1. First-order indices

The 74 first-order Sobol' indices were estimated for each QoI for different values of the sample size: $n = 10^3, 10^4, 10^5, 10^6$. For each index and each value of n , both its estimate and a 95% confidence interval are provided. These results are plotted on Figure 2 for output Y_{01} , namely the maximum surface chlorophyll concentration. As expected, the size of the 95% confidence interval decreases to zero as n increases, and $n = 10^6$ seems to be a sufficiently large sample size to get accurate estimates (size of the 95% confidence interval smaller than 0.01), which allows a clear ranking of the largest indices.

Only ten parameters have a first-order index greater than 0.01 (their values are reported in Table 4). Quite similar results are actually obtained for the other QoIs, and it appears that only 15 or 9 model parameters have a first-order Sobol' index greater than 0.01 for at least 1 or 6 QoIs, respectively (the

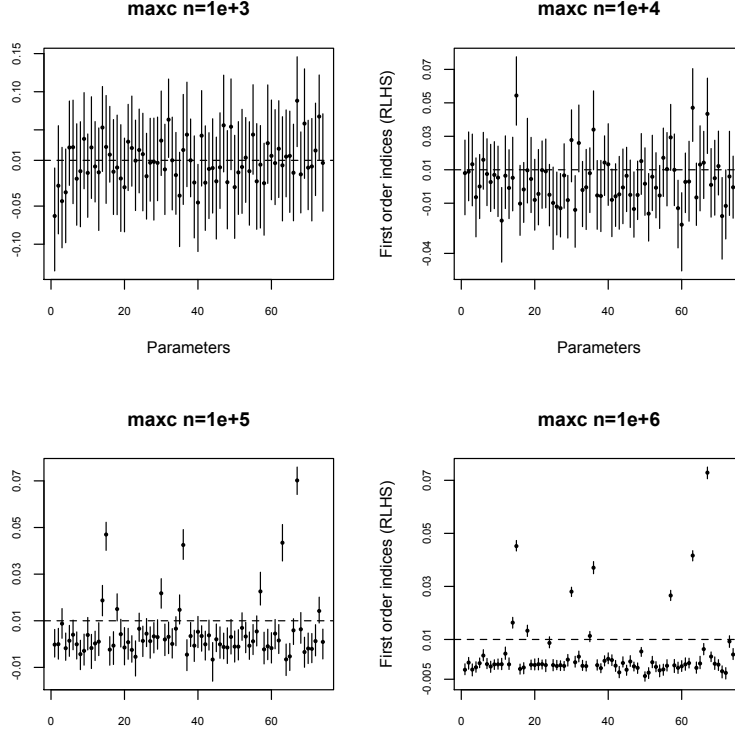


Figure 2: Estimated first-order indices (y -axis) with their 95% confidence interval for the 74 model parameters (x -axis), for $n = 10^3, 10^4, 10^5$ and 10^6 , in the case of the output Y_{01} . The dashed horizontal line corresponds to a threshold arbitrarily chosen to be 0.01. Confidence intervals were obtained with a bootstrap procedure (e.g. Archer et al. (1997)) and a bootstrap sample size of 100. Note that Sobol' estimates are not constrained to be positive, which explains that parts of confidence intervals may be below 0.

361 threshold 0.01 was chosen to guarantee a clear separation between inputs with
 362 high and low first-order indices). This is summarized in Figure 3, where these
 363 most influential model parameters are clearly visible. This mostly highlights
 364 the important sensitivity of our QoIs to the parameterization of excretion for
 365 bacteria, of grazing and ingestion for mesozooplankton, and of the variation
 366 of light limitation for phytoplankton.

367 This small number of influential parameters may indicate that an efficient
 368 reduction of this model could be performed to produce a reduced order
 369 model involving many fewer parameters. Note however that the sum of
 370 first-order indices is equal to 0.371. This is far less than 1, which clearly

371 indicates that the outputs considered are not simple additive models of the
372 form $Y = f_1(X_1) + \dots + f_d(X_d)$. There are some interaction effects of the
373 parameters on the QoIs, which motivates the investigation of second-order
374 interaction effects.

parameter number (see Table C.9)	67	15	63	36	30	57	14	18	35
estimated index	0.0729	0.0452	0.0416	0.0370	0.0280	0.0266	0.0163	0.0133	0.0113
estimated error	0.0022	0.0020	0.0019	0.0023	0.0018	0.0019	0.0020	0.0021	0.0019

Table 4: Estimation of first-order Sobol’ indices for the output Y_{01} (annual maximum of chlorophyll concentration). The estimated error is the radius (half of the length) of the 95% confidence interval.

375

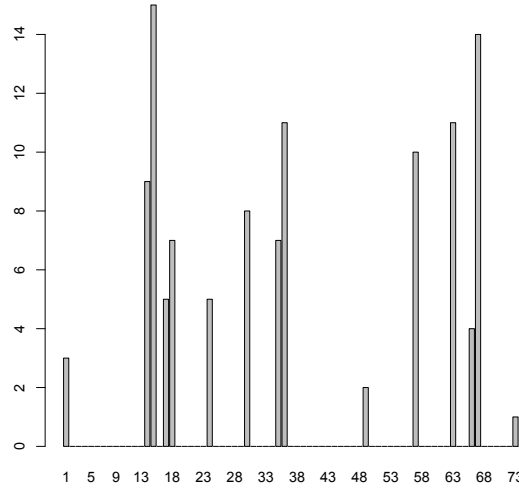


Figure 3: Number of quantities of interest (y -axis) for which the estimate of the first-order index is greater than 0.01. The 15 parameters are: 1 14 15 17 18 24 30 35 36 49 57 63 66 67 73 (x -axis).

376 5.2. Second order indices

377 As indicated before, the `sobolroalhs` function allows the estimation of
378 closed second-order indices at a cost of $2n$ model evaluations. Let us recall
379 that the closed second-order index corresponding to the j^{th} and k^{th} param-
380 eters is defined as:

$$S_{\{j,k\}}^{\text{closed}} = S_{\{j\}} + S_{\{k\}} + S_{\{j,k\}} \quad (9)$$

and corresponds to the sum of the main and interaction effects due to these two parameters.

A specificity for the estimation of closed second-order indices with the replication procedure is that n has to be chosen equal to q^2 , where $q \geq d - 1$ is a prime number denoting the number of levels of the orthogonal array (see Appendix A for more details). A value of $q = 227$, i.e. $2n = 103\,058$, was necessary to achieve sufficiently accurate estimates of the indices.

Estimates of the $d(d-1)/2 = 2701$ closed second-order indices, along with their corresponding 95% bootstrap confidence intervals, were then computed for each QoI. They are displayed in Figure 4 for the output Y_{01} (annual maximum of chlorophyll concentration, which was already chosen in Figure 2 and Table 4). It clearly appears that very few of these 2701 indices are significant. Given the definition (9) of the $S_{\{j,k\}}^{\text{closed}}$, most of them obviously correspond to at least one influential parameter listed in Table 4, as clearly displayed in Figure 5a.

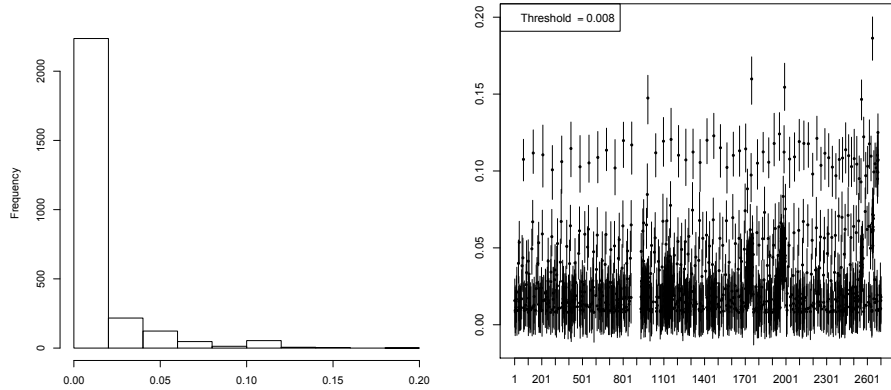


Figure 4: Estimation of second-order closed indices for QoI Y_{01} . Left panel: histogram of the values of the 2701 indices. Right panel: values of these indices with their associated 95% bootstrap confidence intervals (with a bootstrap sample size equal to 100). For sake of clarity, only values greater than 0.008 are shown.

Using these closed second-order index estimates and the previous estimates of first-order index, we compute consistent estimates of unclosed second-order Sobol' indices $S_{\{j,k\}}$ with a bootstrap algorithm. These indices quantify the fraction of variance of the QoI due to these two parameters that cannot be explained by the sum of the main effect of X_i and X_j , but

401 only by their interaction. They are displayed in Figure 5b in the particular
 402 case of output Y_{01} . New interesting information arises from this plot. First,
 403 it appears that input parameters with strong direct influence (i.e. strong
 404 first order indices) may or may not have significant interactions with other
 405 parameters. For instance, parameters 36 and 57 (see Table C.9), related
 406 to semi-saturation for ingestion by MesZ, and fraction of grazing used for
 407 growth of MesZ, have weak second order indices. Conversely, parameters
 408 15 and 18, related to variation of light limitation for NanP and to optimal
 409 temperature for NanP, which already correspond to large first order indices,
 410 have also strong interactions with many other parameters (see Figure 6a).
 411 Secondly, some other parameters that have not been noticed yet, since they
 412 have weak first order indices, show strong interactions with numerous other
 413 parameters. Let us mention in particular parameters 8 and 12, related to
 414 NH_4 semisaturation for PicP and to optimal PAR for NanP (see Figure 6b).
 415 These results are summarized in Figure 6 (recall that the value 0.01 for the
 416 threshold in Figure 6 was chosen to guarantee a clear separation between
 417 inputs with high and small first-order indices).

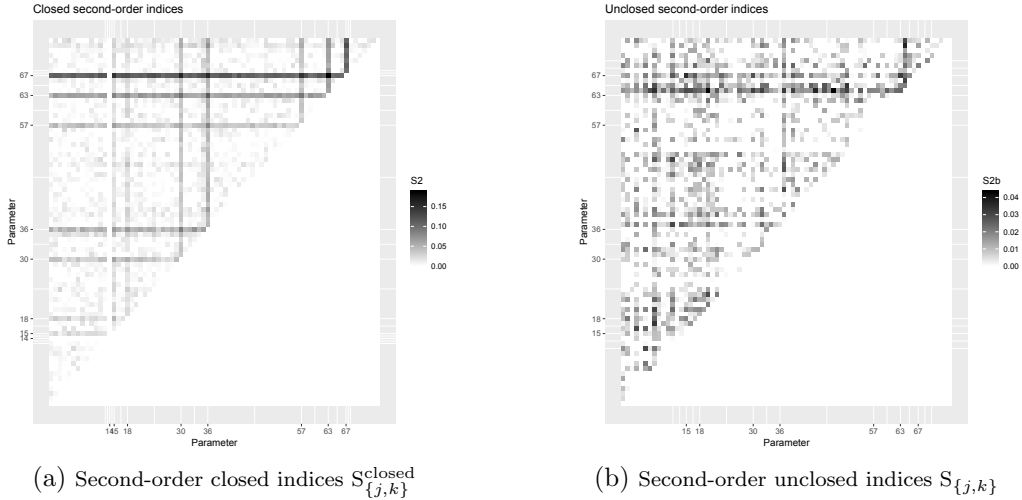


Figure 5: Maps (74×74) of the second-order closed and unclosed Sobol indices for QoI Y_{01} . The x and y axes correspond to the number of the parameters, and the grey scale to the value of the index. Note that the numbers indicated on the axes correspond to parameters with high first-order indices.

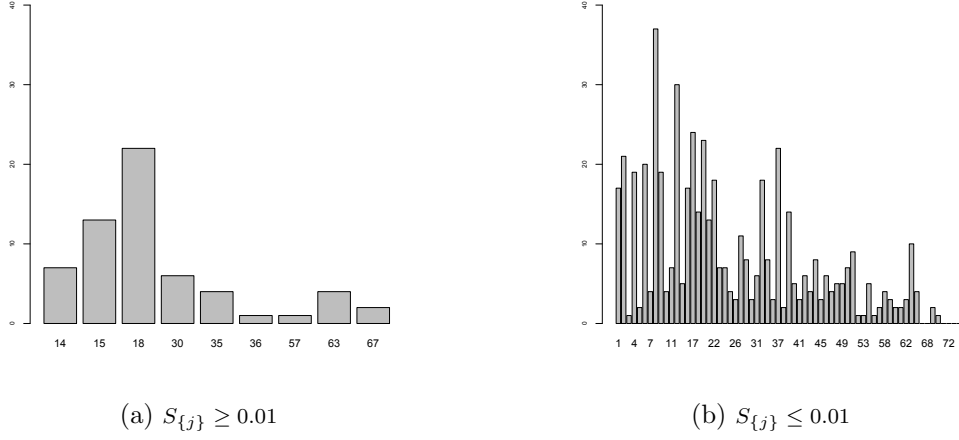


Figure 6: Numbers of second-order unclosed indices $S_{\{j,k\}}$ ($j \neq k$) greater than 0.008, as a function of j , for QoI Y_{01} . Left panel: input parameters with large first-order indices ($S_{\{j\}} \geq 0.01$). Right panel: input parameters with weak first-order indices ($S_{\{j\}} < 0.01$).

At this stage, the information from the first- and second-order indices remains incomplete. Higher order interactions can be expected, especially for highly parameterized models like marine biogeochemistry. A natural step forward in the study is then to compute total indices.

5.3. Total order indices

The total Sobol' index $S_{\{j\}}^{\text{tot}}$ ($j \in \{1, \dots, d\}$) defined in Equation (6), expresses the overall sensitivity of some QoI to the input X_j . As mentioned in Section 2 and Appendix A, a competitive procedure to estimate simultaneously all first and total Sobol' indices was introduced in Saltelli (2002) and implemented by the function `sobolSalt` of the R package `sensitivity`. Its cost is, however, still quite high since it requires $(d + 2)n$ model evaluations. However, this linear computational complexity with regard to the input space dimension d cannot be avoided. Therefore, depending on the computational resources available for the study, computing such indices is not always affordable, although they bring important information since input parameters with a very low value for their total Sobol' index can be fixed to a nominal value in a calibration procedure.

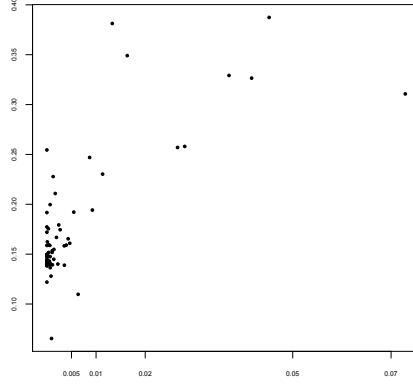


Figure 7: Scatterplot of first-order (x -axis) and total (y -axis) Sobol' indices for output Y_{01} . Note that, for sake of readability, the x -axis is stretched with regard to the y -axis.

435 A scatter plot of first-order and total indices is displayed in Figure 7.
 436 We observe that the parameters contribute to the total variance primarily
 437 through their interactions with other parameters (the dots are far above the
 438 line $y = x$). However, the relation between first-order and total indices is not
 439 linear. More precisely, parameters having a large main effect generally also
 440 contribute to the total variance through their interactions with other param-
 441 eters (Figure 8a). Nevertheless, the relative importance of these interactions
 442 may vary significantly. For instance, parameter 67 (temperature variation
 443 of excretion for bacteria), which corresponds to the largest first-order index,
 444 has a total index smaller than several other parameters with much smaller
 445 first-order indices. On the other hand, some parameters with a very small
 446 main effect make a non-negligible contribution to the total variance via their
 447 interactions with other parameters (Figure 8b). This heterogeneous distri-
 448 bution of the relative importance of the first-order effect with regard to the
 449 total effect is summarized in Figure 9, where the ratio $S_{\{j\}}/S_{\{j\}}^{\text{tot}}$ is displayed.
 450

451 5.4. Comparison with a local analysis

452 The majority of previous sensitivity analyses performed on ocean biogeo-
 453 chemistry models have used local methods, such as the gradient method or
 454 methods where parameters are varied one-at-a-time. It is therefore inter-
 455 esting to compare the results we have just obtained using the global Sobol'
 456 indices method with equivalent results obtained using local methods applied

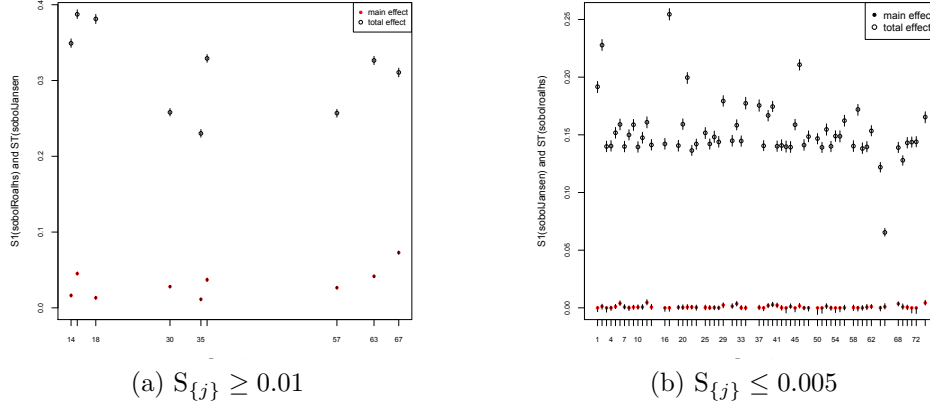


Figure 8: y -axis: values of the first-order and total Sobol' indices for the 74 parameters (x -axis), and corresponding confidence intervals, for output Y_{01} . Left panel: parameters with first-order index larger than 0.01 only. Right panel: parameters with first-order index smaller than 0.005 only.

457 to the same MODECOGeL model.

458 We have thus computed the gradients of our QoI with regard to the 74
 459 input parameters. The partial derivatives are approximated by

$$\frac{\partial Y}{\partial X_j}(x_1, \dots, x_d) \simeq \frac{Y(x_1, \dots, x_{j-1}, x_j + \alpha, x_{j+1}, \dots, x_d) - Y(x_1, \dots, x_d)}{\alpha} \quad (10)$$

460 where α is a small value ensuring very good accuracy of the estimate (in our
 461 case, we checked that any value of α between $10^{-3} \sqrt{\text{Var}(X_j)}$ and $10^{-6} \sqrt{\text{Var}(X_j)}$
 462 is relevant and leads to the same value for the gradient). $d + 1$ (i.e. 75
 463 in our case) model runs are required. The local character of this gradi-
 464 ent appears clearly in (10): the gradient is computed for a particular value
 465 $(X_1, \dots, X_d) = (x_1, \dots, x_d)$. In our case, we choose to compute it for the
 466 mean values of the 74 parameters (listed in Table C.9), i.e. for $(x_1, \dots, x_d) =$
 467 $(E(X_1), \dots, E(X_d))$.

468 As can be seen in Figure 10a and in Table 5, according to this gradient,
 469 the two parameters 67 and 66 (temperature variation of excretion for bac-
 470 teria and mesozooplankton respectively) have a very strong impact on Y_{01} .
 471 The importance of parameter 67 was already noted previously, since it corre-
 472 sponds to the largest first-order Sobol' index, and to one of the largest total
 473 indices. However, parameter 66 does not appear as one of the most influen-

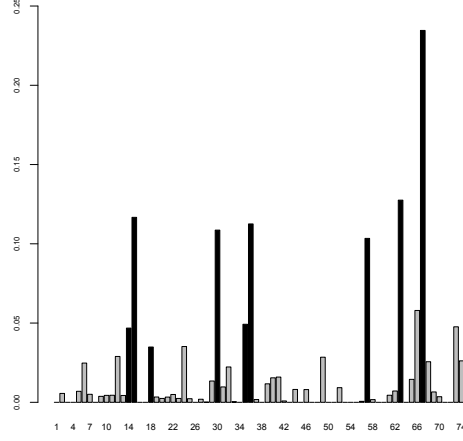


Figure 9: Estimation of the ratio $S_{\{j\}}/S_{\{j\}}^{\text{tot}}$ (y -axis) for $j = 1, \dots, 74$ (x -axis). Black bars correspond to parameters with first-order index larger than 0.01.

474 tial parameters from the global SA: its ranking is just 12th among first-order
 475 indices, and 73rd of 74 considering total indices. Conversely several param-
 476 eters that were clearly highlighted by the global SA do not correspond to
 477 large values of the derivative. This is the case, for instance, for parameter 14
 478 (variation of light limitation for picophytoplankton): its ranking is 3rd con-
 479 sidering total indices and 7th considering first-order indices, but only 21st
 480 considering derivatives. Similar remarks can be made for parameters 30, 36
 481 and 57 (see Table 6). In the same way, the local analysis indicates that
 482 parameters 15 and 18 (variation of light limitation and optimal temperature
 483 for nanophytoplankton) are ten times less important than parameters 66 and
 484 67. In contrast, the total Sobol’ indices show that they are actually more
 485 important (Table 5).

486 In summary, the local analysis identifies parameters 66 and 67 as impor-
 487 tant, and only to a much lesser extent a few others. However, the comparison
 488 with Sobol’ indices clearly shows that this is only local information. Rely-
 489 ing only on this local information would lead us to waste effort tuning some
 490 parameters and to neglect entirely other important ones.

A complementary way to look at this local gradient is to compute

$$S_{\{j\}}^{\text{loc}} = \frac{\text{Var}(X_j)}{\text{Var}(Y)} \left(\frac{\partial Y}{\partial X_j} \right)^2$$

491 This local index is non dimensional. It can be interpreted as a “relative gradi-

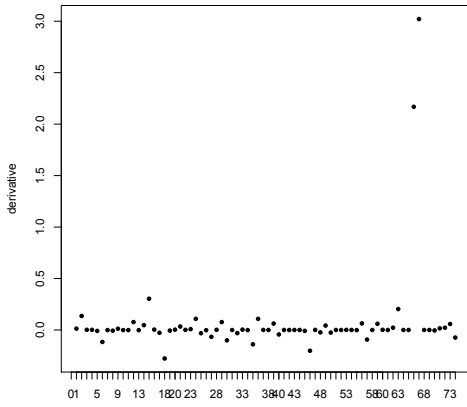
j	67	66	15	18	63	46
$\partial Y/\partial X_j$	3.02	2.17	0.30	-0.28	0.20	-0.20
$S_{\{j\}}$	0.073	0.006	0.045	0.013	0.042	0.002
$S_{\{j\}}^{\text{tot}}$	0.311	0.110	0.387	0.381	0.326	0.211

Table 5: Largest values of the local derivative $\partial Y/\partial X_j$ and corresponding first-order and total Sobol' indices $S_{\{j\}}$ and $S_{\{j\}}^{\text{tot}}$

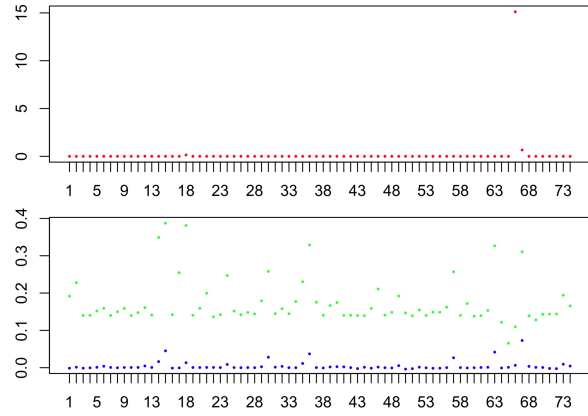
j	2	14	15	18	30	35	36	46	57	63	66	67
$\partial Y/\partial X_j$	8 th		3 rd	4 th		7 th		6 th		5 th	2 nd	1 st
$S_{\{j\}}$		7 th	2 nd	8 th	5 th		4 th		6 th	3 rd		1 st
$S_{\{j\}}^{\text{tot}}$		3 rd	1 st	2 nd	7 th		4 th		8 th	5 th		6 th

Table 6: Top eight ranking of the local derivative $\partial Y/\partial X_j$, and first-order and total Sobol' indices $S_{\{j\}}$ and $S_{\{j\}}^{\text{tot}}$

ent", since it could also be written as $(\partial Y'/\partial X'_j)^2$, where $Y' = Y/\sqrt{\text{Var}(Y)}$ and $X'_j = X_j/\sqrt{\text{Var}(X_j)}$ are normalized versions of Y and X_j . Its interest for our study lies in the fact that, in the case of a linear dependency $Y = a_0 + a_1 X_1 + \dots + a_d X_d$ and of the independence of input parameters, this index is equal to the first order Sobol' index. As can be seen in Figure 10b, this index indicates that the most (and almost only) influential parameter is parameter 66 (and, to a much lesser extent, parameter 67). As noted previously, this result is contradicted by the global SA. This disagreement is a clear illustration of the strongly nonlinear character of the MODECOGeL model.



(a) Derivative



(b) $S_{\{j\}}^{\text{loc}}$ derivative (upper panel) and corresponding first order and total Sobol' indices (lower panel)

Figure 10: Derivative $\frac{\partial Y_{01}}{\partial X_i}$ (left), non dimensional derivative $S_{\{j\}}^{\text{loc}} = \frac{\text{Var}(X_i)}{\text{Var}(Y_{01})} \left(\frac{\partial Y_{01}}{\partial X_i} \right)^2$ (right, upper panel), and first order and total Sobol' indices (right, lower panel) as functions of the number of the parameter (x -axis). The derivatives are computed for $(x_1, \dots, x_d) = (E(X_1), \dots, E(X_d))$.

5.5. Computational costs

The relative computational costs of the different algorithms are displayed in Table 7, along with the estimated error related to the Sobol' indices (i.e. the half-length of the 95% bootstrap confidence interval). As mentioned earlier, the computation of total indices requires a number of model runs proportional to $d + 2$, which is much more than for the evaluation of only first- and second-order indices. Moreover, the estimation error is inversely proportional to \sqrt{n} , as is mainly the case in Monte Carlo methods, consistent with the rate of convergence in the central limit theorem. Note that other designs of experiments, such as quasi Monte Carlo methods, are not considered in this paper, since we do not assume any mathematical regularity in the underlying model. In the present case, given the order of magnitude of the indices, $n = 10^5$ appears to be a good compromise between accuracy and computational cost.

Estimation of Sobol' indices	sobolSalt n=10 ⁵		roalhs n=10 ³	roalhs n=10 ⁴	roalhs n=10 ⁵	roalhs n=10 ⁶	roalhs q=227
	S1	ST	S1	S1	S1	S1	S2
Estimated Error							
<i>maximum value</i>	0.0076	0.0064	0.077	0.025	0.0077	0.0025	0.024
<i>mean value</i>	0.0065	0.0047	0.062	0.020	0.0064	0.0020	0.018
<i>standard deviation</i>	2.0 10 ⁻⁷	2.3 10 ⁻⁷	2.7 10 ⁻⁵	3.4 10 ⁻⁶	5.7 10 ⁻⁸	3.2 10 ⁻⁸	2.7 10 ⁻⁷
Number of evaluations	7.6 10 ⁶		2 10 ³	2 10 ⁴	2 10 ⁵	2 10 ⁶	$q^2 \simeq 10^5$

Table 7: Statistics (maximum and mean values, standard deviation) related to the estimated error over all 74 parameters, and number of model runs required for the estimation of the Sobol' indices.

515

6. Conclusion

The added value of SA is primarily to help scientists understanding the behavior of their models, and identifying their most influential parameters. Then, as a by-product in the context of model tuning, it helps focusing optimization effort on these few parameters (see for instance the two review papers by Oschlies (2006) and Dowd et al. (2014), and references herein, on data assimilation in the context of ocean biogeochemical models).

In the present work, a global sensitivity analysis based on Sobol' indices is presented in the context of a realistic ocean biogeochemical model. Without delving into complex physical interpretations, the aim of this study is

526 to demonstrate the potential interest of this systematic and mathematically
 527 sound approach for such applications. This method quantifies the influence
 528 of uncertain input parameters on an arbitrary number of output quantities
 529 of interest, either through their direct effect or through their mutual in-
 530 teractions. In the present study, which was performed with a full Monte
 531 Carlo method, thus avoiding any restrictive hypothesis on the effective di-
 532 mension of the model or on its regularity, we started by computing first- and
 533 second-order indices. This highlighted on the one hand the strong influence
 534 of some input parameters and of some second-order interactions, and on the
 535 other that some higher order interactions were probably significant. Total
 536 Sobol' indices were thus computed, which provided complementary informa-
 537 tion. Note that total indices can also be a helpful source of information for
 538 constructing reduced models, since they identify parameters whose values
 539 could be frozen. These results were then compared to those derived using a
 540 local OAT gradient-based SA, where each parameter was varied individually
 541 around a point in parameter space corresponding to the average. It clearly
 542 illustrated the fact that a local SA can lead us to focus our attention on non
 543 globally relevant parameters and to neglect the influence of globally impor-
 544 tant ones. This comparison also brought out the strongly nonlinear character
 545 of the biogeochemical model considered here.

546 More generally, from a biogeochemical point of view, the methods de-
 547 scribed in this paper can provide a substantial help in the design of the
 548 models, in the tuning of their complexity, and in the understanding of their
 549 dependence to uncertainties in the parameters. It was shown indeed in our
 550 test application that the traditional sensitivity studies based on local meth-
 551 ods can lead to misleading conclusions, and thus to a misunderstanding of
 552 the influence of every individual process on the key results of the simula-
 553 tions. Accounting for the nonlinear effects and for the interactions between
 554 the various processes is needed to decide what must be explicitly resolved
 555 by the model (and carefully tuned) and what can be roughly parameterized.
 556 Until recently, this ambition was very difficult to achieve. The take home
 557 message for further ocean biogeochemical research is that, due to the con-
 558 stant increase of computational resources, combined with recent advances in
 559 estimation procedures for the sensitivity measures, it is now possible to per-
 560 form a global sensitivity analysis based on the computation of Sobol' indices.
 561 Large sets of input parameters can be considered and indices as expensive as
 562 total indices can be feasibly computed by taking advantage of grid computing
 563 environments which scale easily to high dimensional applications.

564 Appendix A. Estimation procedure for Sobol' indices

565 In this paper, we used a Monte Carlo based procedure for the estimation
 566 of Sobol' indices (Sobol', 1993), which we describe in detail below. Let us
 567 first introduce some notation. Let \mathbf{u} be a non-empty subset of $\{1, \dots, d\}$. Let
 568 $\mathbf{X}^1 = (X_1^1, \dots, X_d^1)$ and $\mathbf{X}^2 = (X_1^2, \dots, X_d^2)$ be two independent copies of
 569 the input vector $\mathbf{X} = (X_1, \dots, X_d)$. Let $(\mathbf{X}^{1,i})_{i=1, \dots, n}$ be an *independent and*
 570 *identically distributed* (i.i.d.) sample of size n of vector \mathbf{X}^1 and $(\mathbf{X}^{2,i})_{i=1, \dots, n}$
 571 be an i.i.d. sample of size n of vector \mathbf{X}^2 . For any $\mathbf{u} \subset \{1, \dots, d\}$, we define
 572 the d -dimensional vector $\mathbf{X}_{\mathbf{u}} = (X_{\mathbf{u},1}, \dots, X_{\mathbf{u},d})$ in the following way: for any
 573 $j \in \mathbf{u}$, $X_{\mathbf{u},j}$ is equal to X_j^1 , and for any $j \in \mathbf{u}^c$, $X_{\mathbf{u},j} = X_j^2$. The n -sample
 574 $(\mathbf{X}_{\mathbf{u}}^i)_{i=1, \dots, n}$ is defined in a similar manner.

575 And we define the corresponding output variables:

$$Y = f(\mathbf{X}^1), Y_{\mathbf{u}} = f(\mathbf{X}_{\mathbf{u}}) \text{ and } Y^i = f(\mathbf{X}^{1,i}), Y_{\mathbf{u}}^i = f(\mathbf{X}_{\mathbf{u}}^i). \quad (\text{A.1})$$

576 It is possible to prove that $S_{\mathbf{u}}^{\text{closed}} = \frac{\text{Cov}(Y, Y_{\mathbf{u}})}{\text{Var}[Y]}$ (see, e.g., Lemma 1.2 in
 577 Janon et al. (2014)). Based on the above formula, we propose the following
 578 estimator for $S_{\mathbf{u}}^{\text{closed}}$:

$$\hat{S}_{\mathbf{u},n}^{\text{closed}} = \frac{\frac{1}{n} \sum_{i=1}^n Y^i Y_{\mathbf{u}}^i - \left(\frac{1}{n} \sum_{i=1}^n \frac{Y^i + Y_{\mathbf{u}}^i}{2} \right)^2}{\frac{1}{n} \sum_{i=1}^n \frac{(Y^i)^2 + (Y_{\mathbf{u}}^i)^2}{2} - \left(\frac{1}{n} \sum_{i=1}^n \frac{Y^i + Y_{\mathbf{u}}^i}{2} \right)^2}. \quad (\text{A.2})$$

579 This estimator was first introduced in Monod et al. (2006). Its asymptotic
 580 properties are stated in Janon et al. (2014, Propositions 2.2 and 2.5). See
 581 also Gamboa et al. (2016) for further asymptotic and non-asymptotic results.

582 Since $S_{\{j\}} = S_{\{j\}}^{\text{closed}}$ (see Eq. (5)), we therefore estimate $S_{\{j\}}$ with $\hat{S}_{\{j\},n} =$
 583 $\hat{S}_{\{j\},n}^{\text{closed}}$.

584 Recall now the law of total variance which can be found, e.g., in (Weiss,
 585 2006, pages 385-386):

$$\text{Var}[Y] = \text{Var}(\mathbb{E}(Y|\mathbf{X}_{\mathbf{u}})) + \mathbb{E}(\text{Var}(Y|\mathbf{X}_{\mathbf{u}})). \quad (\text{A.3})$$

586 From the definition of total Sobol' indices, we can prove that:

$$S_{\{j\}}^{\text{tot}} = \frac{\mathbb{E}(\text{Var}(Y|X_{\{j\}^c}))}{\text{Var}[Y]}. \quad (\text{A.4})$$

From (A.3) and (A.4) we deduce $S_{\{j\}}^{\text{tot}} = 1 - S_{\{j\}^c}^{\text{closed}}$. We thus define the estimator of $S_{\{j\}}^{\text{tot}}$ as:

$$\hat{S}_{\{j\},n}^{\text{tot}} = 1 - \hat{S}_{\{j\}^c,n}^{\text{closed}}.$$

587 These estimators require a large number of model evaluations. Indeed,
 588 the estimation of all first-order indices actually requires $(d+1)n$ evaluations
 589 of the model: $Y^i, Y_{\{j\}}^i, i = 1, \dots, n, j = 1, \dots, d$. The estimation of all closed
 590 second-order indices requires $(\frac{d(d-1)}{2} + 1)n$ model evaluations of the model:
 591 $Y^i, Y_{\{j,k\}}^i, i = 1, \dots, n, j \neq k = 1, \dots, d$.

592 The Monte Carlo sample size n is directly related to the accuracy of the
 593 estimation via the rate of convergence in the central limit theorem, which is
 594 of order \sqrt{n} (see Proposition 2.2 in Janon et al. (2014)).

595 To circumvent this linear or quadratic dependence of the cost (in terms
 596 of number of model evaluations) in the input space dimension d , the authors
 597 in Tissot and Prieur (2015) proposed a procedure based on replicated Latin
 598 Hypercube Sampling (LHS) (e.g. Lemieux, 2009) or replicated randomized
 599 orthogonal arrays to estimate all first-order or all closed second-order) Sobol'
 600 indices respectively. The ideas in Tissot and Prieur (2015) generalize some pi-
 601 oneer work in Mara and Rakoto Joseph (2008). We refer to Tissot and Prieur
 602 (2015) for a detailed description of the procedure, and a rigorous analysis of
 603 its asymptotic properties. In the present paper, we applied the replication
 604 procedure to estimate all first-order Sobol' indices in Subsection 5.1 and to
 605 estimate all closed second-order Sobol' indices in Subsection 5.2. More pre-
 606 cisely, the replication procedure in Subsection 5.1 allows us to estimate all
 607 the first-order Sobol' indices with only two replicated d -dimensional LHS of
 608 size n , that is with only $2n$ model evaluations. The replication procedure in
 609 Subsection 5.2 allows to estimate all closed second-order Sobol' indices with
 610 only two replicated d -dimensional randomized orthogonal arrays of strength
 611 two and size n , that is with only $2n$ model evaluations. Due to constraints
 612 in the construction of orthogonal arrays of strength two, n must be chosen
 613 as q^2 , with q a prime number greater or equal to $d-1$. Note that orthogonal
 614 arrays were introduced for the first time in Kishen (1942).

615 If higher order interactions are expected in the model (of any order greater
 616 or equal to three), one may be interested in estimating total Sobol' indices.
 617 The replication procedure does not adapt to the estimation of total Sobol'
 618 indices, mainly because of the constraints in the construction of orthogonal
 619 arrays of strength higher or equal to three. In Saltelli (2002), the authors

620 propose two different procedures, both based on combinatorial tricks: the
621 first one allows the estimation of all first-order and total Sobol' indices with
622 a cost of $(d + 2)n$ model evaluations (see Theorem 1 in Saltelli (2002), the
623 second one leads to a double estimate of all first-order, closed second-order
624 and total Sobol' indices at a cost of $(2d + 2)n$ model evaluations (see Theorem
625 2 in Saltelli (2002)). To our knowledge, the procedure in (Saltelli, 2002,
626 Theorem 1) is the most competitive one if the estimation of total Sobol'
627 indices is involved (see also Gilquin et al. (2017)). It is the one we have
628 applied in Subsection 5.3.

629 Appendix B. Description of biogemical fluxes

630 The biogeochemical fluxes ($C_j \rightarrow C_i$) parameterized in MODECOGeL are
631 summarized in Table B.8. Each flux depends on several parameters, which
632 are indicated by referring to the parameter list in Table C.9 in Appendix
633 C. To give an idea of the role of each parameter in MODECOGeL, the
634 biogeochemical fluxes are organized into several categories using different
635 colors:

- 636 • **Primary production (green)** is the growth of phytoplankton by pho-
637 tosynthesis. In Table B.8, this corresponds to all fluxes from nutrients
638 (C_1, C_2) to phytoplankton (C_3, C_4, C_5). Parameters govern the maxi-
639 mum growth rate (1–3), and the dependence on nutrient concentrations
640 (4–10), to solar irradiance (11–16), and to temperature (18–20).
- 641 • **Secondary production (blue)** is the growth of zooplankton by graz-
642 ing of phytoplankton or by assimilation of bacteria and particulate
643 organic matters. In Table B.8, this corresponds to all fluxes to zoo-
644 plankton (C_6, C_7, C_8). Parameters govern the ingestion rate (28–33),
645 the dependence on prey concentration (34–36), the efficiency according
646 to the type of prey (37–43), and the fraction actually used for growth
647 (55–59).
- 648 • **Mortality (red)** of living species, including a parameterization of pre-
649 dation by higher trophic levels. In Table B.8, this corresponds to all
650 fluxes from phytoplankton or zooplankton or bacteria (C_3 to C_9) to
651 particulate organic matter (C_{11} or C_{12}). Parameters govern mortality
652 rates (44–50) and predation (51–53).
- 653 • **Exudation (magenta)** by phytoplankton. In Table B.8, this corre-
654 sponds to all fluxes from phytoplankton (C_3, C_4, C_5) to dissolved or-
655 ganic nitrogen (C_{10}). Parameters are exudation rates (25–27).
- 656 • **Excretion (pink)** by zooplankton and bacteria. In Table B.8, this
657 corresponds to all fluxes from zooplankton or bacteria (C_6 to C_9) to
658 ammonium and dissolved organic nitrogen (C_2 and C_{10}). Parameters
659 govern excretion rates (60–63), the dependence on temperature (64–
660 67), the tradeoff between ammonium and dissolved organic matter (68),
661 and the excreted fraction of predation (54).

- 662 • **Growth of bacteria (yellow)** from ammonium and dissolved organic
663 matter. In Table B.8, this corresponds to all fluxes to bacteria (C_9).
664 Parameters govern the growth rate (23–24).
- 665 • **Decomposition of particulate organic matter (orange)**. In Ta-
666 ble B.8, this corresponds to all fluxes from particulate organic matter
667 (C_{11} or C_{12}) to dissolved organic nitrogen (C_{10}). Parameters are de-
668 composition rates (69–70).
- 669 • **Nitrification (brown)**. In Table B.8, this corresponds to the flux
670 from ammonium (C_2) to nitrate (C_1). The parameter is the nitrification
671 rate (72).

	Nutrients		Phytoplanktons			Zooplanktons			BAC	DON & POM		
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}
C_1			1,4,5, 11,14, 17,20	2,4,6, 12,15, 18,21	3,4,7, 13,16, 19,22							
C_2	72	72	1,8, 11,14, 17,20	2,9, 12,15, 18,21	3,10, 13,16, 19,22				23,24			
C_3			1,5,8, 11,14, 17,20, 25,44			28,31, 34,55				25	44	
C_4				2,6,9, 12,15, 18,21, 26,45			29,32, 35,56			26	45	
C_5					3,7,10, 13,16, 19,22, 27,46			30,33, 36,37, 57		27	46	
C_6						28,31, 34,47, 55,60, 64	29,32, 35,39, 56			60,64, 68	47	
C_7							29,32, 35,48, 56,61, 65	30,33, 36,40, 57		61,65, 68	48	
C_8								30,33, 36,49, 52,53, 54,57, 62,66		54,62, 66,68		49,51, 52,53
C_9						28,31, 34,38, 55	29,32, 35,56		23,24, 50,63, 67		50	
C_{10}									23,24			
C_{11}							29,32, 35,41, 58	30,33, 36,42, 59		69	69	
C_{12}								30,33, 36,43, 59		70		70

Table B.8: Biogeochemical fluxes from variable C_i (line i) to variable C_j (column j). Numbers in the boxes refer to parameter indices, given in Table C.9.

672 Appendix C. Model parameters

673 This table describes the different probability distributions chosen for in-
674 put parameters.

Index	Name	Unit	Pdf	Mean	Std	Std/Mean
1	PicP max growth rate	t^{-1}	$\Gamma(25, 0.12)$	3.	0.6	20%
2	NanP max growth rate	t^{-1}	$\Gamma(25, 0.1)$	2.5	0.5	20%
3	MicP max growth rate	t^{-1}	$\Gamma(25, 0.08)$	2.	0.4	20%
4	dependence of NO3 limitation to NH4	C^{-1}	$\Gamma(400, 0.00365)$	1.46	0.073	5%
5	NO3 semisaturation for PicP	C	$\Gamma(4, 0.125)$	0.5	0.25	50%
6	NO3 semisaturation for NanP	C	$\Gamma(4, 0.175)$	0.7	0.35	50%
7	NO3 semisaturation for MicP	C	$\Gamma(4, 0.25)$	1.0	0.5	50%
8	NH4 semisaturation for PicP	C	$\Gamma(4, 0.075)$	0.3	0.15	50%
9	NH4 semisaturation for NanP	C	$\Gamma(4, 0.125)$	0.5	0.25	50%
10	NH4 semisaturation for MicP	C	$\Gamma(4, 0.175)$	0.7	0.35	50%
11	optimal PAR for PicP	I	$\Gamma(25, 0.4)$	10.	2.	20%
12	optimal PAR for NanP	I	$\Gamma(25, 0.6)$	15.	3.	20%
13	optimal PAR for MicP	I	$\Gamma(25, 0.8)$	20.	4.	20%
14	variation of light limitation for PicP	—	$-\Gamma(4, 0.2)$	-0.8	0.4	50%
15	variation of light limitation for NanP	—	$-\Gamma(4, 0.175)$	-0.7	0.35	50%
16	variation of light limitation for MicP	—	$-\Gamma(4, 0.15)$	-0.6	0.3	50%
17	optimal temperature for PicP	T	$\mathcal{N}(15, 3^2)$	15.	3.	20%
18	optimal temperature for NanP	T	$\mathcal{N}(15, 3^2)$	15.	3.	20%
19	optimal temperature for MicP	T	$\mathcal{N}(16, 3.2^2)$	16.	3.2	20%
20	variation of temp. limitation for PicP	—	$-\Gamma(4, 0.125)$	-0.5	0.25	50%
21	variation of temp. limitation for NanP	—	$-\Gamma(4, 0.125)$	-0.5	0.25	50%
22	variation of temp. limitation for MicP	—	$-\Gamma(4, 0.1375)$	-0.55	0.275	50%
23	bacteria growth limitation	—	$\Gamma(4, 0.15)$	0.6	0.3	50%
24	semisaturation for BAC growth	C	$\Gamma(4, 0.125)$	0.5	0.25	50%
25	exudation ratio for PicP	—	$\Gamma(4, 0.015)$	0.06	0.03	50%
26	exudation ratio for NanP	—	$\Gamma(4, 0.0125)$	0.05	0.025	50%
27	exudation ratio for MicP	—	$\Gamma(4, 0.01)$	0.04	0.02	50%
28	max ingestion rate for NanZ	t^{-1}	$\Gamma(25, 0.12)$	3.	0.6	20%
29	max ingestion rate for MicZ	t^{-1}	$\Gamma(25, 0.08)$	2.	0.4	20%
30	max ingestion rate for MesZ	t^{-1}	$\Gamma(25, 0.06)$	1.5	0.3	20%
31	threshold ingestion for NanZ	C	$\Gamma(4, 0.0125)$	0.05	0.025	50%
32	threshold ingestion for MicZ	C	$\Gamma(4, 0.0075)$	0.03	0.015	50%
33	threshold ingestion for MesZ	C	$\Gamma(4, 0.0025)$	0.01	0.005	50%
34	semisaturation for ingestion by NanZ	C	$\Gamma(4, 0.125)$	0.5	0.25	50%
35	semisaturation for ingestion by MicZ	C	$\Gamma(4, 0.1875)$	0.75	0.375	50%
36	semisaturation for ingestion by MesZ	C	$\Gamma(4, 0.25)$	1.	0.5	50%
37	efficiency of MesZ on MicP	—	$\beta(4.2, 1.05)$	0.8	0.16	20%
38	efficiency of NanZ on BAC	—	$\beta(4.2, 1.05)$	0.8	0.16	20%
39	efficiency of MicZ on NanZ	—	$\beta(4.2, 1.05)$	0.8	0.16	20%
40	efficiency of MesZ on MicZ	—	$\beta(4.2, 1.05)$	0.8	0.16	20%
41	efficiency of MicZ on MOP1	—	$\beta(19.8, 79.2)$	0.2	0.04	20%
42	efficiency of MesZ on MOP1	—	$\beta(19.8, 79.2)$	0.2	0.04	20%
43	efficiency of MesZ on MOP2	—	$\beta(19.8, 79.2)$	0.2	0.04	20%
44	mortality rate for PicP	t^{-1}	$\Gamma(4, 0.015)$	0.06	0.03	50%
45	mortality rate for NanP	t^{-1}	$\Gamma(4, 0.0125)$	0.05	0.025	50%
46	mortality rate for MicP	t^{-1}	$\Gamma(4, 0.01)$	0.04	0.02	50%
47	mortality rate for NanZ	t^{-1}	$\Gamma(4, 0.015)$	0.06	0.03	50%

Index	Name	Unit	Pdf	Mean	Std	Std/Mean
48	mortality rate for MicZ	t^{-1}	$\Gamma(4, 0.0125)$	0.05	0.025	50%
49	mortality rate for MesZ	t^{-1}	$\Gamma(4, 0.0075)$	0.03	0.015	50%
50	mortality rate for BAC	t^{-1}	$\Gamma(4, 0.015)$	0.06	0.03	50%
51	threshold for predation	C	$\Gamma(4, 0.005)$	0.02	0.01	50%
52	maximum predation rate on MesZ	t^{-1}	$\Gamma(4, 0.25)$	1.	0.5	50%
53	semisaturation for predation on MesZ	C	$\Gamma(4, 0.25)$	1.	0.5	50%
54	excreted fraction of predation on MesZ	—	$\beta(2.33, 4.67)$	0.333	0.167	50%
55	fraction of grazing used for growth of NanZ	—	$\beta(4.2, 1.05)$	0.8	0.16	20%
56	fraction of grazing used for growth of MicZ	—	$\beta(4.2, 1.05)$	0.8	0.16	20%
57	fraction of grazing used for growth of MesZ	—	$\beta(4.2, 1.05)$	0.8	0.16	20%
58	fraction of POM used for growth of MicZ	—	$\beta(12, 12)$	0.5	0.1	20%
59	fraction of POM used for growth of MesZ	—	$\beta(12, 12)$	0.5	0.1	20%
60	excretion rate for NanZ	t^{-1}	$\Gamma(4, 0.0375)$	0.15	0.075	50%
61	excretion rate for MicZ	t^{-1}	$\Gamma(4, 0.025)$	0.1	0.05	50%
62	excretion rate for MesZ	t^{-1}	$\Gamma(4, 0.0125)$	0.05	0.025	50%
63	excretion rate for BAC	t^{-1}	$\Gamma(4, 0.0375)$	0.15	0.075	50%
64	temperature variation of excretion for NanZ	—	LogGamma	1.05	0.0525	5%
65	temperature variation of excretion for MicZ	—	LogGamma	1.05	0.0525	5%
66	temperature variation of excretion for MesZ	—	LogGamma	1.02	0.051	5%
67	temperature variation of excretion for BAC	—	LogGamma	1.04	0.052	5%
68	fraction of excretion as DOM	—	$\beta(2.75, 8.25)$	0.25	0.125	50%
69	POM1 decomposition rate	t^{-1}	$\Gamma(4, 0.01625)$	0.065	0.0325	50%
70	POM2 decomposition rate	t^{-1}	$\Gamma(4, 0.015)$	0.06	0.03	50%
71	sedimentation velocity for MicP	V	$\Gamma(4, 0.25)$	1.	0.5	50%
72	nitrification rate	t^{-1}	$\Gamma(4, 0.0075)$	0.03	0.015	50%
73	light attenuation coefficient in sea water	—	$\Gamma(25, 0.0016)$	0.04	0.008	20%
74	fraction of photosynthetically active radiation	—	$\Gamma(25, 0.02)$	0.5	0.1	20%

Table C.9: Model parameters X_i . Units are: time t in days, concentration C in mmolN/m³, irradiance I in W/m², and velocity V in m/day. The notation $-\Gamma(.,.)$ means that the parameter is negative and that its opposite follows a $\Gamma(.,.)$ distribution.

References

- G. Archer, A. Saltelli, and I. Sobol'. Sensitivity measures, ANOVA-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(2):99–120, 1997.
- M. Baklouti, V. Faure, L. Pawlowski, and A. Sciandra. Investigation and sensitivity analysis of a mechanistic phytoplankton model implemented in a new modular numerical tool (eco3m) dedicated to biogeochemical modelling. *Progress in Oceanography*, 71(1):34–58, 2006.
- M. Dowd, E. Jones, and J. Parslow. A statistical overview and perspectives on data assimilation for marine biogeochemical models. *Environmetrics*, 25(4):203–213, 2014.

- 686 J.-N. Druon and J. Le Fèvre. Sensitivity of a pelagic ecosystem model to
687 variations of process parameters within a realistic range. *Journal of Marine*
688 *Systems*, 19(1-3):1–26, 1999.
- 689 B. Efron and C. Stein. The jackknife estimate of variance. *The Annals of*
690 *Statistics*, 9(3):586–596, 1981.
- 691 B. Faugeras, M. Lévy, L. Mémer, J. Verron, J. Blum, and I. Charpentier.
692 Can biogeochemical fluxes be recovered from nitrate and chlorophyll data?
693 a case study assimilating data in the northwestern Mediterranean sea at
694 the JGOFS-DYFAMED station. *Journal of Marine Systems*, 40:99–125,
695 2003.
- 696 K. Fennel, M. Losch, J. Schröter, and M. Wenzel. Testing a marine ecosystem
697 model: sensitivity analysis and parameter optimization. *Journal of Marine*
698 *Systems*, 28(1-2):45–63, 2001.
- 699 F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur. Statistical
700 inference for Sobol pick free Monte Carlo method. *Statistics*, 50(4):881–
701 902, 2016.
- 702 L. Gilquin, E. Arnaud, C. Prieur, and A. Janon. Making best use of permuta-
703 tions to compute sensitivity indices with replicated designs. working paper
704 or preprint, June 2017. URL <https://hal.inria.fr/hal-01558915>.
- 705 W. F. Hoeffding. A class of statistics with asymptotically normal distribu-
706 tions. *Annals of Mathematical Statistics*, 19:293–325, 1948.
- 707 T. Homma and A. Saltelli. Importance measures in global sensitivity analysis
708 of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17,
709 1996.
- 710 A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Asymptotic
711 normality and efficiency of two Sobol index estimators. *ESAIM: Probability*
712 *and Statistics*, 18:342–364, 2014.
- 713 K. Kishen. On latin and hyper-graeco-latin cubes and hyper-cubes. *Current*
714 *Science*, 11(3):98–99, 1942.
- 715 I. Kriest, A. Oschlies, and S. Khatriwala. Sensitivity analysis of simple global
716 marine biogeochemical models. *Global Biogeochemical Cycles*, 26(2), 2012.

- 717 G. Lacroix. *Simulation de l'écosystème pélagique de la mer Ligure à l'aide*
718 *d'un modèle unidimensionnel. Etude du bilan de matière et de la variabilité*
719 *saisonnnière, interannuelle et spatiale.* PhD thesis, Université Pierre et
720 Marie Curie, Paris, France, 1998.
- 721 G. Lacroix and M. Grégoire. Revisited ecosystem model (MODECOGeL) of
722 the Ligurian sea: seasonal and interannual variability due to atmospheric
723 forcing. *J. Marine Syst.*, 37(4):229–258, 2002.
- 724 G. Lacroix and P. Nival. Influence of meteorological variability on primary
725 production dynamics in the Ligurian Sea (NW Mediterranean sea) with a
726 1D hydrodynamic/biological model. *J. Marine Syst.*, 16(1-2):23–50, 1998.
- 727 C. Lemieux. *Monte Carlo and quasi-Monte Carlo sampling.* Springer Science
728 & Business Media, 2009.
- 729 T. Mara and O. Rakoto Joseph. Comparison of some efficient methods to
730 evaluate the main effect of computer model factors. *Journal of Statistical*
731 *Computation and Simulation*, 78(2):167–178, 2008.
- 732 J.C. Marty and J. Chiavérini. Hydrological changes in the ligurian sea (nw
733 mediterranean, dyfamed site) during 19952007 and biogeochemical conse-
734 quences. *Biogeosciences*, 7(7):2117–2128, 2010.
- 735 H. Monod, C. Naud, and D. Makowski. Uncertainty and sensitivity analysis
736 for crop models. In D. Wallach, D. Makowski, and J.W. Jones, editors,
737 *Working with Dynamic Crop Models: Evaluation, Analysis, Parameteri-*
738 *zation, and Applications*, chapter 4, pages 55–99. Elsevier, 2006.
- 739 D.J. Morris, D.C. Speirs, A.I. Cameron, and M.R. Heath. Global sensitivity
740 analysis of an end-to-end marine ecosystem model of the north sea: Factors
741 affecting the biomass of fish and benthos. *Ecological Modelling*, 273:251–
742 263, 2014.
- 743 M.D. Morris. Factorial sampling plans for preliminary computational exper-
744 iments. *Technometrics*, 33(2):161–174, 1991.
- 745 J. C. J. Nihoul and S. Djenidi. Perspectives in three-dimensional modelling
746 of the marine system. In J.C.J. Nihoul and B.M. Jamart, editors, *Three-*
747 *Dimensional Models of Marine and Estuarine Dynamics*, pages 1–34. El-
748 sevier, Amsterdam, 1987.

- 749 M. Omlin, R. Brun, and P. Reichert. Biogeochemical model of lake Zürich:
750 sensitivity, identifiability and uncertainty analysis. *Ecological Modelling*,
751 141(1-3):105–123, 2001.
- 752 A. Oschlies. On the use of data assimilation in biogeochemical modelling. In
753 *Ocean Weather Forecasting*, pages 525–547. Springer, 2006.
- 754 A. B. Owen. Orthogonal arrays for computer experiments, integration and
755 visualization. *Statistica Sinica*, pages 439–452, 1992.
- 756 G. Pujol, B. Iooss, A. Janon, P. Lemaitre, L. Gilquin, L. Le Gratiet,
757 T. Touati, B. Ramos, J. Fruth, and S. De Veiga. *Sensitivity: Global Sen-
758 sitivity Analysis of Model Outputs*, 2017. R package version 1.15.0.
- 759 A. Saltelli. Making best use of model evaluations to compute sensitivity
760 indices. *Computer Physics Communications*, 145(2):280–297, 2002.
- 761 A. Saltelli, K. Chan, and E.M. Scott. *Sensitivity analysis*. Wiley Series in
762 Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2000.
- 763 S. Sankar, L. Polimene, L. Marin, N.N. Menon, A. Samuelsen, R. Pastres,
764 and S. Ciavatta. Sensitivity of the simulated oxygen minimum zone to bio-
765 geochemical processes at an oligotrophic site in the arabian sea. *Ecological
766 modelling*, 372:12–23, 2018.
- 767 M. Schartau, P. Wallhead, J. Hemmings, U. Löptien, I. Kriest, S. Krishna,
768 B.A. Ward, T. Slawig, and A. Oschlies. Reviews and syntheses: parameter
769 identification in marine planktonic ecosystem modelling. *Biogeosciences
770 (BG)*, 14(6):1647–1701, 2017.
- 771 V. Scott, H. Kettle, and C. Merchant. Sensitivity analysis of an ocean carbon
772 cycle model in the north Atlantic: an investigation of parameters affecting
773 the air-sea CO₂ flux, primary production and export of detritus. *Ocean
774 Science*, 7(3):405–419, 2011.
- 775 I. Sobol’. Sensitivity analysis for nonlinear mathematical models. *Mathemat-
776 ical Modeling and Computational Experiment*, 1:407–414, 1993.
- 777 J.-Y. Tissot and C. Prieur. A randomized orthogonal array-based procedure
778 for the estimation of first- and second-order Sobol’ indices. *Journal of
779 Statistical Computation and Simulation*, 85(7):1358–1381, 2015.

- 780 J.F. Tjiputra, D. Polzin, and A. Winguth. Assimilation of seasonal chloro-
781 phyll and nutrient data into an adjoint three-dimensional ocean carbon
782 cycle model: sensitivity analysis and ecosystem parameter optimization.
783 *Global biogeochemical cycles*, 21(1), 2007.
- 784 S. Wang, N. Flipo, and T. Romary. Time-dependent global sensitivity anal-
785 ysis of the c-rive biogeochemical model in contrasted hydrological and
786 trophic contexts. *Water research*, 144:341–355, 2018.
- 787 N. A. Weiss. *A course in probability*. Addison-Wesley, Boston, MA, 2006.
788 URL <http://cds.cern.ch/record/735662>.